



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number: 0 075 444 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification: 23.12.92 (51) Int. Cl⁵: C12N 15/10, C07H 21/04,

(21) Application number: 82304880.6

C12N 1/20, //C12R1/19,
C12P21/02

(22) Date of filing: 16.09.82

Divisional application 87202220 filed on 13.11.87.

(54) Methods and products for facile microbial expression of DNA sequences.

(30) Priority: 18.09.81 US 303687

structure of mRNA and efficiency of translation initiation" & GENE 1980, 9(1-2), 1-12,

(43) Date of publication of application:
30.03.83 Bulletin 83/13

(73) Proprietor: GENENTECH, INC.
460 Point San Bruno Boulevard
South San Francisco California 94080(US)

(45) Publication of the grant of the patent:
23.12.92 Bulletin 92/52

(72) Inventor: De Boer, Herman Albert
1035 San Carlos Avenue

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

EI Granada California 94018(US)
Inventor: Seeburg, Peter H.

(56) References cited:

800 Kirkham
San Francisco California 94122(US)

EP-A- 0 001 931
EP-A- 0 022 242
GB-A- 2 073 245

Inventor: Heyneker, Herbert L.
2621 Easton Drive
Burlingame California 94010(US)

PROC. NATL. ACAD. SCI. USA, vol. 74, no. 10,
October 1977, pages 4163-4167; D.A.

(74) Representative: Armitage, Ian Michael et al
MEWBURN ELLIS & CO. 2/3 Cursitor Street
London EC4A 1BQ(GB)

STEEGE: "5'-Terminal nucleotide sequence
of Escherichia coli lactose repressor mRNA:
features of translational initiation and
reinitiation sites"

CHEMICAL ABSTRACTS, vol. 92, no. 23, 9th
June 1980, page 201, no. 192883j, Columbus,
Ohio, US; D. ISERENTANT et al.: "Secondary

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

Description

The present invention provides methods and means for preparing DNA sequences that provide messenger RNA having improved translation characteristics. The resulting messenger RNA may be highly efficient in translation to give substantial amounts of polypeptide product that is normally heterologous to the host microorganism. The DNA sequences which are ultimately expressed, that is, transcribed into messenger RNA (mRNA) which is in turn translated into polypeptide product, are, in essential part, synthetically prepared, in accordance with this invention, utilizing means that favor the substantial reduction or elimination of secondary and/or tertiary structure in the corresponding transcribed mRNA. An absence or 5 substantial reduction in such secondary/tertiary structure involving the 5' end of mRNA permits effective 10 recognition and binding of ribosomes(s) to the mRNA for subsequent translation. Thus, the efficiency of translation is not hindered or impaired by conformational impediments in the structure of the transcribed mRNA. Methods and means for measuring mRNA secondary/tertiary structure are also described as well as associated means designed to insure that secondary/tertiary structure is kept below certain preferred limits.

15 This invention is exemplified by the preparation of various preferred protein products.

With the advent of recombinant DNA technology, the controlled microbial production of an enormous variety of useful polypeptides has become possible, putting within reach the microbially directed manufacture of hormones, enzymes, antibodies, and vaccines useful against a wide variety of diseases. Many mammalian polypeptides, such as human growth hormone and leukocyte interferons, have already been 20 produced by various microorganisms.

One basic element of recombinant DNA technology is the plasmid, an extrachromosomal loop of double-stranded DNA found in bacteria oftentimes in multiple copies per cell. Included in the information encoded in the plasmid DNA is that required to reproduce the plasmid in daughter cells (i.e., a "replicon") and ordinarily, one or more selection characteristics, such as resistance to antibiotics, which permit clones 25 of the host cell containing the plasmid of interest to be recognized and preferentially grown in selective media. The utility of such bacterial plasmids lies in the fact that they can be specifically cleaved by one or another restriction endonuclease or "restriction enzyme", each of which recognizes a different site on the plasmidic DNA. Heterologous genes or gene fragments may be inserted into the plasmid by endwise joining at the cleavage site or at reconstructed ends adjacent to the cleavage site. (As used herein, the term 30 "heterologous" refers to a gene not ordinarily found in, or a polypeptide sequence ordinarily not produced by, a given microorganism, whereas the term "homologous" refers to a gene or polypeptide which is found in, or produced by the corresponding wild-type microorganism.) Thus formed are so-called replicable expression vehicles.

DNA recombination is performed outside the microorganism, and the resulting "recombinant" plasmid 35 can be introduced into microorganisms by a process known as transformation and large quantities of the heterologous gene-containing recombinant plasmid are obtained by growing the transformant. Moreover, where the gene is properly inserted with reference to portions of the plasmid which govern the transcription and translation of the encoding DNA, the resulting plasmid can be used to actually produce the polypeptide sequence for which the inserted gene codes, a process referred to as expression. Plasmids which express 40 a (heterologous) gene are referred to as replicable expression vehicles.

Expression is initiated in a DNA region known as the promotor. In some cases, as in the lac and trp systems discussed infra, promotor regions are overlapped by "operator" regions to form a combined promotor-operator. Operators are DNA sequences which are recognized by so-called repressor proteins which serve to regulate the frequency of transcription initiation from a particular promoter. In the transcription 45 phase of expression, RNA polymerase recognizes certain sequences in and binds to the promoter DNA. The binding interaction causes an unwinding of the DNA in this region, exposing the DNA as a template for synthesis of messenger RNA. The messenger RNA serves as a template for ribosomes which bind to the messenger RNA and translate the mRNA into a polypeptide chain having the amino acid sequence for which the RNA/DNA codes. Each amino acid is encoded by a nucleotide triplet or "codon" which 50 collectively make up the "structural gene", i.e., that part of the DNA sequence which encodes the amino acid sequence of the expressed polypeptide product.

After binding to the promoter, RNA polymerase initiates the transcription of DNA encoding a ribosome binding site including a translation initiation or "start" signal (ordinarily ATG, which in the resulting messenger RNA becomes AUG), followed by DNA sequences encoding the structural gene itself. So-called 55 translational stop codons are transcribed at the end of the structural gene whereafter the polymerase may form an additional sequence of messenger RNA which, because of the presence of the translational stop signal, will remain untranslated by the ribosomes.

Ribosomes bind to the binding site provided on the messenger RNA, in bacteria ordinarily as the mRNA is

being formed, and direct subsequently the production of the encoded polypeptide, beginning at the translation start signal and ending at the previously mentioned stop signal(s). The resulting product may be obtained by lysing the host cell and recovering the product by appropriate purification from other bacterial proteins.

5 Polypeptides expressed through the use of recombinant DNA technology may be entirely heterologous, functional proteins, as in the case of the direct expression of human growth hormone, or alternatively may comprise a bioactive heterologous polypeptide portion and, fused thereto, a portion of the amino acid sequence of a homologous polypeptide, as in the case of the production of intermediates for somatostatin and the components of human insulin. In the latter cases, for example, the fused homologous polypeptide
10 comprised a portion of the amino acid sequence for beta galactosidase. In those cases, the intended bioactive product is rendered bioinactive within the fused, homologous/ heterologous polypeptide until it is cleaved in an extracellular environment. Fusion proteins like those just mentioned can be designed so as to permit highly specific cleavage of the precursor protein from the intended product, as by the action of cyanogen bromide on methionine, or alternatively by enzymatic cleavage. See, eg., G.B. Patent Publication
15 No. 2 007 676 A.

If recombinant DNA technology is to fully sustain its promise, systems must be devised which optimize expression of gene inserts, so that the intended polypeptide products can be made available in controlled environments and in high yields.

20 Promoter Systems

As examples, the beta lactamase and lactose promoter systems have been advantageously used to initiate and sustain microbial production of heterologous polypeptides. Details relating to the make-up and construction of these promoter systems have been published by Chang et al., Nature 275, 617 (1978) and
25 Itakura et al., Science 198, 1056 (1977), which are hereby incorporated by reference. More recently, a system based upon tryptophan, the so-called trp promoter system, has been developed. Details relating to the make-up and construction of this system have been published by Goeddel et al., Nucleic Acids Research 8, 4057 (1980) and Kleid et al., U.S.S.N. 133, 296, filed March 24, 1980, (or the equivalent European Patent Publication 0036776) which are hereby incorporated by reference. Numerous other
30 microbial promoters have been discovered and utilized and details concerning their nucleotide sequences, enabling a skilled worker to ligate them functionally within plasmid vectors, have been published -- see, e.g. Siebenlist et al., Cell 20, 269 (1980), which is incorporated herein by this reference.

Historically, recombinant cloning vehicles (extrachromosomal duplex DNA having, inter alia., a functional origin of replication) have been prepared and used to transform microorganisms -- cf. Ullrich et al., Science 196, 1313 (1977). Later, there were attempts to actually express DNA gene inserts encoding a heterologous polypeptide. Itakura et al. (Science 198, 1056 (1977)) expressed the gene encoding somatostatin in E. coli. Other like successes followed, the gene inserts being constructed by organic synthesis using newly refined technology. In order, among other things, to avoid possible proteolytic degradation of the polypeptide product within the microbe, the genes were ligated to DNA sequences coding for a precursor polypeptide.
40 Extracellular cleavage yielded the intended protein product, as discussed above.

In the case of larger proteins, chemical synthesis of the underlying DNA sequence proved unwieldy. Accordingly, resort was had to the preparation of gene sequences by reverse transcription from corresponding messenger RNA obtained from requisite tissues and/or culture cells. These methods did not always prove satisfactory owing to the termination of transcription short of the entire sequence; and/or the desired
45 sequence would be accompanied by naturally occurring precursor leader or signal DNA. Thus, these attempts often have resulted in incomplete protein product and/or protein product in non-cleavable conjugate form -- cf. Villa-Komaroff et al., Proc. Natl. Acad. Sci. (USA) 75, 3727 (1978) and Seeburg et al., Nature 276, 795 (1978).

In order to avoid these difficulties, Goeddel et al., Nature 281, 544 (1979), constructed DNA, inter alia
50 encoding human growth hormone, using chemically synthesized DNA in conjunction with enzymatically synthesized DNA. This discovery thus made available the means enabling the microbial expression of hybrid DNA (combination of chemically synthesized DNA with enzymatically synthesized DNA), notably coding for proteins of limited availability which would probably otherwise not have been produced economically. The hybrid DNA (encoding heterologous polypeptide) is provided in substantial portion, preferably a majority, via reverse transcription of mRNA, while the remainder is provided via chemical synthesis. In a preferred embodiment, synthetic DNA encoding the first 24 amino acids of human growth hormone (HGH) was constructed according to a plan which incorporated an endonuclease restriction site in the DNA corresponding to HGH amino acids 23 and 24. This was done to facilitate a connection with

downstream HGH cDNA sequences. The various 12 oligonucleotide long fragments making up the synthetic part of the DNA were chosen following then known criteria for gene synthesis: avoidance of undue complementarity of the fragments, one with another, except, of course, those destined to occupy opposing sections of the double stranded sequence; avoidance of AT rich regions to minimize transcription 5 termination; and choice of "microbially preferred codons." Following synthesis, the fragments were permitted to effect complementary hydrogen bonding and were ligated according to methods known per se. This work is described in published British Patent Specification 2055382 A, which corresponds to Goeddel et al., U.S.S.N. 55126, filed July 5, 1979 which is hereby incorporated by this reference.

While the successful preparation and expression of such hybrid DNA provided a useful means for 10 preparing heterologous polypeptides, it did not address the general problem that eucaryotic genes are not always recognized by procaryotic expression machinery in a way which provides copious amounts of end product. Evolution has incorporated sophistication unique to discrete organisms. Bearing in mind that the eucaryotic gene insert is heterologous to the procarytic organism, the relative inefficiency in expression often observed can be true for any gene insert whether it is produced chemically, from cDNA or as a 15 hybrid. Thus, the criteria used to construct the synthetic part of the gene for HGH, defined above, are not the sole factors influencing expression levels. For example, concentrating on codon choice as the previous workers have done--cf. British Patent Specification 2007676 A -- has not been completely successful in raising the efficiency of expression towards maximal expression levels.

Guarante et al., Science 209, 1428 (1980) experimented with several hybrid ribosome binding sites, 20 designed to match the number of base pairs between the Shine-Dalgarno sequence and the ATG of some known E. coli binding sites, their work suggesting that the reason(s) for observed relatively low efficiencies of eucaryotic gene expression by procaryote organisms is more subtle.

That the initiation of mRNA translation may be a multicomponent process is illustrated by work reported by Iserentant and Fiers, Gene 9, 1 (1980). They postulate that secondary structure of mRNA is one of the 25 components influencing translation efficiency and imply that the initiation codon and ribosome interaction site of secondary structured, folded mRNA must be "accessible." However, what those workers apparently mean by "accessible" is that the codon and site referred to be located on the loop, rather than the stem, of the secondary structure models they have hypothesized. Shine et al, Nature, 285 (1980), 456 and Bahramian, J. theor. Biol., both emphasise the seeming importance of secondary structure in mRNA to 30 achieve efficient translation.

The present invention is based upon the discovery that the presence of secondary/tertiary conformational structure in the mRNA interferes with the initiation and maintenance of ribosomal binding during the translation phase of heterologous gene expression.

The present invention, relating to these findings, uniquely provides methods and means for providing 35 efficient expression of heterologous gene inserts by the requisite microbial host. The present invention is further directed to a method of microbi ally producing heterologous polypeptides, utilizing specifically tailored heterologous gene inserts in microbial expression vehicles, as well as associated means. It is particularly directed to the use of synthetically derived gene insert portions that are prepared so as to both encode the desired polypeptide product and provide mRNA that has minimal secondary/ tertiary structure 40 and hence is accessible for efficient ribosomal translation.

In preferred embodiments of the present invention, synthetic DNA is provided for a substantial portion of the initial coding sequence of a heterologous gene insert, and optionally, upstream therefrom through the ATG translational start codon and ribosome binding site. The critical portion of DNA is chemically synthesized, keeping in mind two factors: 1) the creation of a sequence that codes for the initial (N-terminal) 45 amino acid sequence of a polypeptide comprising a functional protein or bioactive portion thereof and 2) the assurance that said sequence provides, on transcription, messenger RNA that has a secondary/tertiary conformational structure which is insufficient to interfere with its accessibility for efficient ribosomal translation, as herein defined. Such chemical synthesis may use standard organic synthesis using modified mononucleotides as building blocks such as according to the method of Crea et al., Nucleic Acids Research 8, 2331 (1980) and/or the use of site directed mutagenesis of DNA fragments such as according to the 50 method of Razin et al., Proc. Natl. Acad Sci (USA) 75, 4268 (1978) and/or synthetic primers on certain appropriately sequenced DNA fragments followed by specific cleavage of the desired region.

The present invention is directed to a process of preparing DNA sequences comprising nucleotides arranged sequentially so as to encode the proper amino acid sequence of a given polypeptide.

55 This method may involve obtaining a substantial portion of the DNA coding sequence of a given polypeptide via means other than chemical synthesis, most often by reverse transcription from requisite tissue and/or culture cell messenger RNA. This fragment encodes the C-terminal portion of the polypeptide and is ligated, in accordance herewith, to a remainder of the coding sequence, e.g. obtained by chemical

synthesis, optionally including properly positioned translational start and stop signals and upstream DNA through the ribosome binding site and the first nucleotide (+1) of the resultant messenger RNA. The synthetic fragment is designed by nucleotide choice dependent on conformation of the corresponding messenger RNA according to the criteria as herein discussed.

5 The thus prepared DNA sequences are suited for insertion and use in replicable expression vehicles designed to direct the production of the heterologous polypeptide in a transformant microorganism. In these vehicles, the DNA sequence is operably linked to promotor systems which control its expression. The invention is further directed to the replicable expression vehicles and the transformant microorganisms so produced as well as to cultures of these microorganisms in fermentation media. This invention is further 10 directed to associated methods and means and to specific embodiments for the directed production of messenger RNA transcripts that are accessible for efficient ribosomal translation.

15 Excluded from the present invention, for example, is the hybrid DNA encoding human growth hormone (HGH) as disclosed by Goeddel et al., *Nature* 281, 544 (1979). While this particular hybrid DNA was successfully expressed to produce the intended product, the concept of the present invention was not appreciated by these workers (and hence not taught by them) and consequently was not practised in the fortuitous preparation of their expressible hybrid DNA for HGH. This hybrid DNA has the following sequence (Table 1):

20

25

30

35

40

45

50

55

Table I

5 1
 met phe pro thr ile pro leu ser arg leu phe asp asn ala met
 ATG TTC CCA ACT ATA CCA CTA TCT CGT CTA TTC GAT AAC GCT ATG

10 20
 leu arg ala his arg leu his gln leu ala phe asp thr tyr gln
 CTT CGT GCT CAT CGT CTT CAT CAG CTG GCC TTT GAC ACC TAC CAG

15 40
 glu phe glu glu ala tyr ile pro lys glu gln lys tyr ser phe
 GAG TTT GAA GAA GCC TAT ATC CCA AAG GAA CAG AAG TAT TCA TTC

20 60
 leu gln asn pro gln thr ser leu cys phe ser glu ser ile pro
 CTG CAG AAC CUC CAG ACC TCC CTC TCT TTC TCA CAG TCT ATT CCC

25 80
 thr pro ser asn arg glu glu thr gln gln lys ser asn leu glu
 ACA CCC TCC AAC AGG GAG GAA ACA CAA CAG AAA TCC AAC CTA GAG

30 100
 val gln phe leu arg ser val phe ala asn ser leu val tyr gly
 GTG CAG TTC CTC AGG AGT GTC TTC GCC AAC AGC CTA GTG TAC GGC

35 120
 ala ser asp ser asn val tyr asp leu leu lys asp leu glu glu
 GCC TCT GAC AAC GTC TAT GAC CTC CTA AAG GAC CTA GAG GAA

40 140
 gly ile gln thr leu met gly arg leu glu asp gly ser pro arg
 GGC ATC CAA ACG CTG ATG GGG AGG CTG GAA GAT GGC AGC CCC CGG

45 160
 thr gly gln ile phe lys gln thr tyr ser lys phe asp thr asn
 ACT GGG CAG ATC TTC AAG CAG ACC TAC AGC AAG TTC GAC ACA AAC

50 180 191
 ser his asn asp asp ala leu leu lys asn tyr gly leu leu tyr
 TCA CAC AAC GAT GAC GCA CTA CTC AAG AAC TAC GGG CTG CTC TAC

55 cys phe arg lys asp met asp lys val glu thr phe leu arg ile
 TGC TTC AGG AAG GAC ATG GAC AAG GTC GAG ACA TTC CTG CGC ATC

The chemically synthetic DNA sequences hereof extend preferably from the ATG translation initiation site, and optionally upstream therefrom a given distance, to or beyond the transcription initiation site (labelled +1 by convention), and to sequences downstream encoding a substantial part of the desired polypeptide. By way of preference, the synthetic DNA comprises upwards of approximately 75 or more nucleotide pairs of the structural gene representing about the proximal (N-terminal) 25 amino acids of the intended polypeptide. In particularly preferred embodiments, the synthetic DNA sequence extends from about the translation initiation site (ATG) to about nucleotide 75 of the heterologous gene. In alternative terms, the synthetic DNA sequence comprises nucleotide pairs from +1 (transcription initiation) to about nucleotide 100 of the transcript.

Because of the degeneracy of the genetic code, there is substantial freedom in codon choice for any given amino acid sequence. Given this freedom, the number of different DNA nucleotide sequences encoding any given amino acid sequence is exceedingly large, for example, upwards of 2.6×10^5

possibilities for somatostatin consisting of only 14 amino acids. Again, the present invention provides methods and means for selecting certain of these DNA sequences, those which will efficiently prepare functional product. For a given polypeptide product hereof, the present invention provides means to select, from among the large number possible, those DNA sequences that provide transcripts, the conformational structure of which admits of accessibility for operable and efficient ribosomal translation.

Conformational structure of mRNA transcripts is a consequence of hydrogen bonding between complementary nucleotide sequences that may be separated one from another by a sequence of noncomplementary nucleotides. Such bonding is commonly referred to as secondary structure. So-called tertiary structures may add to the conformation of the overall molecule. These structures are believed to be a result of spatial interactions within one or more portions of the molecule -- so-called stacking interactions. In any event, the conformational structure of a given mRNA molecule can be determined and measured. Furthermore, we have now discovered that certain levels of conformational structure of mRNA transcripts interfere with efficient protein synthesis, thus effectively blocking the initiation and/or continuation of translation (elongation) into polypeptide product. Accordingly, levels at which such conformational structure does not occur, or at least is minimal, can be predicted. Nucleotide choice can be prescribed on the basis of the predictable, permissible levels of conformational structure, and preferred gene sequences determined accordingly.

The measurement of mRNA conformational structure is determined, in accordance herewith, by measuring the energy levels associated with the conformational structure of the mRNA molecule.

In determining such energy levels, the thermodynamic disassociation energy connected with one or a series of homologous base pairings is calculated, for example according to the rules of Tinoco et al., *Nature New Biol* 246, 40(1973). In the calculation used herein (not that of Tinoco et al, *supra*), AT base pairing is assigned an associated energy level of about -1.2 Kcal/mole while a CG base pairing is assigned an associated energy level of about -2 Kcal/mole. Adjacent homologous pairings are more than additive, doubtless due to stacking interactions and other associative factors. In any event, it has been determined that in those instances where, according to this calculation regional base pairing interactions result in energy levels stronger than about -12 kcal/mole (that is, values expressed arithmetically in numbers less than about -12 kcal/mole) for a given homologous sequence, such interactions are likely sufficient to hinder or block the translation phase of expression, most probably by interfering with accessibility for necessary ribosomal binding.

A given DNA sequence is screened as follows: A first series of base pairs, e.g., approximately the first six base pairs, are compared for homology with the corresponding reverse last base pairs of the gene. If such homology is found, the associate energy levels are calculated according to the above considerations. The first series of base pairs is next compared with the corresponding last base pairs up to the penultimate base pair of the gene and the associative energy levels of any homology calculated. In succession the first series of base pairs is next compared with the corresponding number of base pairs up to the antepenultimate base pair, and so on until the entire gene sequence is compared, back to front. Next, the series of base pairs beginning one downstream from the first series, e.g. base pairs 2 to 7 of the prior example, is compared with the corresponding number from the end and progressively toward the front of the gene, as described above. This procedure is repeated until each base pair is compared for homology with all other regions of the gene and associated energy levels are determined. Thus, for example in Figure 3 there are provided results of such scanning and calculating for two genes - those encoding natural bovine growth hormone (BGH) and synthetic (i.e., hybrid) BGH. It can be seen that natural BGH contains two regions of homology considered relevant herein (i.e., according to this calculation, energy level greater than about -12 kcal/mole), to wit, six base pairs from base pair 33 to 38 with homologous pairs 96 to 101 and six base pairs from 46 to 51 with 73 to 78. The first is not significant for present purpose, despite the energy level (-15.40 kcal/mole), presumably because the region of homology lies downstream a sufficient distance so as not to be influential on translation efficiency. The second region is significant as evidenced by the poor yields of product as described herein cf. infra. The synthetic BGH gene where such region of homology was eliminated provided good yields of intended protein.

An embodiment of the present invention will now be described by way of example with reference to the accompanying drawings, in which:

Figure 1 depicts the amino acid and nucleotide sequences of the proximal portions of natural BGH, synthetic HGH, and synthetic BGH. The amino acids and nucleotides in natural BGH that are different from those in synthetic HGH are underlined. The nucleotides in the proximal portion of the synthetic BGH gene that differ from those in the natural BGH gene also are underlined. The position of the PVUII restriction site at the end of the proximal portion of these genes is indicated.

In arriving at the synthetic BGH gene encoding the proper amino acid sequence for BGH, the

nucleotide sequences of natural BGH and synthetic HGH were compared. Nucleotide selections were made based upon the synthetic HGH gene for construction of the synthetic BGH gene taking into account also the latitude permitted by the degeneracy of the genetic code, using a minimum of nucleotide changes from the synthetic HGH sequence.

5 Figure 2 depicts the nucleotide sequences of the sense strands of both natural and synthetic BGH genes along with the transcribed portions of the respective preceding trp-promotor sequences. The first nucleotide of each transcript is indicated as "+1" and the following nucleotides are numbered sequentially. The sequences are lined up to match the translated coding regions of both genes, beginning at the start codon "ATG" of each (overlined). The transcript of the natural BGH gene shows an area of "secondary 10 structure" due to interactions of nucleotides 46 to 51 with nucleotides 73 to 78, respectively (see Figure 3), thus creating the stem-loop structure depicted. This area is not present in the synthetic BGH gene, removed by virtue of nucleotide changes (see Figure 1), which nevertheless retains the correct amino acid sequence.

15 Figure 3 shows the locations and stabilities of secondary structures in the transcripts of natural and synthetic BGH. (See Figure 2) These locations and stabilities were determined using a nucleotide by nucleotide analysis, as described herein. Each area of significant secondary structure of each proximal portion of gene is listed in the respective table. Thus, for natural BGH versus synthetic BGH, it is noted that the energy levels of "secondary structure" at corresponding portions of the translatable transcripts (namely, nucleotides 46 to 78 comprising a 6 nucleotide long stem in natural BGH versus nucleotides 52 to 84 of 20 synthetic BGH) are markedly different (according to this calculation -15.2 kcal/mole versus greater than -10 kcal/mole), accounting for the observed success of expression of the synthetic BGH gene versus the natural BGH gene, cf. infra. The energy levels indicate the significance of the relative amounts of tolerable "secondary structure", i.e., according to this calculation values arithmetically greater than about -12kcal/mole based upon thermodynamic energy considerations. The significance of location of "secondary 25 structure" can be appreciated by the fact that energy levels calculated for positions 33 to 101 versus 38 to 104 of natural versus synthetic BGH, respectively, did not significantly influence expression levels.

Figure 4 depicts the construction of pBGH 33 used as shown in Figure 5.

Figure 5 depicts the construction of plasmids harboring DNA sequences for hybrid polypeptides: pBHGH 33-1 used as shown in Figure 7, pBHGH being a hybrid of bovine and human growth hormone 30 sequences, and pHBGH a hybrid of human and bovine sequences.

Figure 6 depicts the technique used to assemble the synthetic proximal portion of the BGH gene, pBR 322-01, used in the construction shown in Figure 7.

Figure 7 depicts the construction of the plasmid (pBGH 33-3) harboring the gene for BGH comprising the synthetic proximal portion as shown in Figure 6.

35 Figure 8 depicts the construction of expression plasmid pBGH 33-4 harboring the hybrid BGH gene.

Figure 9 is the result of a polyacrylamide gel segregation of cell protein. Part A shows no BGH production at any cell density using the culture containing natural BGH gene. Part B shows the expression of synthetic BGH gene (lanes BGH #1 and #2) in the same medium as used for Part A. The levels of expression indicated in Part B, as opposed to Part A, reflect the production of BGH in amounts exceeding 40 about 100 thousand copies per cell.

In its most preferred embodiment, the invention is illustrated by the microbial production of bovine growth hormone (BGH). BGH is endogenous in bovine, e.g., cattle, and is responsible for proper physical maturation of the animal. It is also useful for increasing weight gain, feed conversion efficiency, lean to fat ratio, and milk production. Its sequence of 190 amino acids is known. See Dayhoff, Atlas of Protein Sequence and Structure 1972, National Biomedical Research Foundation, Washington, D.C. The present invention makes possible the preparation of commercial quantities of the compound, enabling now its application on a large scale in the animal husbandry industry. An initial approach toward preparing BGH microbially took advantage of a source of bovine pituitary glands. By extraction and purification, the requisite mRNA for BGH was isolated and from it, corresponding cDNA prepared. Thus, this initial work 45 resulted in a gene corresponding, for all intents and purposes, to the natural DNA sequence of BGH. After removal of DNA coding for the presequence and adding a start codon, the cDNA was ligated to a plasmid vector under proper control of a promotor. This plasmid was used to transform E. coli host which was grown under usual conditions. The efficiency of expression of BGH product was poor, a consequence, it was discovered, of conformational structure of the messenger RNA, which greatly reduced its accessibility for 50 ribosomal translation, cf. Figure 3.

For example, it was found that in "natural" BGH mRNA there are regions of complementary homology. One significant region centers around positions +46 to +51 with a homologous region at positions +73 to +78. Secondary structure considerations, in these two defined regions, are thought to create a hairpin

arrangement just downstream from the translation start codon ATG and the ribosome binding site. This conformational arrangement interferes with or prematurely disrupts ribosomal binding, and hence, inhibits translation.

The recognition of this phenomenon prompted investigations into the nature of the DNA sequence in these regions and the discovery of methods and means to obviate the problem. In accordance herewith, advantage was taken of a *Pvu* II endonuclease restriction site at the BGH DNA corresponding to amino acid 24. DNA for the first 24 amino acids of BGH was chemically synthesized, the selection of nucleotides taking into strict account proper coding sequence and resultant mRNA secondary/ tertiary structure considerations. Employing the method defined above, it was found that certain nucleotide base selections would be suitable, on the basis of predicted conformational structure energy levels, to prepare gene sequences properly encoding BGH but devoid of problematic conformational structure. One of these was selected and synthesized. Ligations at the *Pvu* II terminus of the synthetic piece to the cDNA downstream therefrom produced the desired hybrid gene. Construction of a replicable expression vector containing said heterologous gene as an operable insert successfully resulted in efficient expression of BGH in transformed *E. coli* host.

The complete nucleotide (and deduced amino acid) sequence of the thus constructed hybrid BGH gene is as follows:

20

25

30

35

40

45

50

55

1
met phe pro ala met ser leu ser gly leu phe ala asn ala val
ATG TTC CCA GCT ATG TCT CTA TCT GGT CTA TTC GCT AAC GCT GTT

5
leu arg ala gln his leu his gln leu ala ala asp thr phe lys
CTT CGT GCT CAG CAT CTT CAT CAG CTG GCT GAC ACC TTC AAA

10
glu phe glu arg thr tyr ile pro glu gly gln arg tyr ser ile
GAG TTT GAG CGC ACC TAC ATC CCG GAG GGA CAG AGA TAC TCC ATC

15
gln asn thr gln val ala phe cys phe ser glu thr ile pro ala
CAG AAC ACC CAG GTT GCC TTC TGC TTC TCT GAA ACC ATC CCG GCC

20
pro thr gly lys asp glu ala gln gln lys ser asp leu glu leu
CCC ACG GGC AAG GAT GAG GCC CAG AAA TCA GAC TTG GAG CTG

25
leu arg ile ser leu leu leu ile gln ser trp leu gly pro leu
CTT CGC ATC TCA CTG CTC ATC CAG TCG TGG CTT GGG CCC CTG

30
gln phe leu ser arg val phe thr asn ser leu val phe gly thr
CAG TTC CTC AGC AGA GTC TTC ACC AAC AGC TCG GTG TTT GGC ACC

35
ser asp arg val tyr glu lys leu lys asp leu glu glu gly ile
TCG GAC CGT GTC TAT GAG AAG CTG AAG GAC CTG GAG GAA GGC ATC

40
leu ala leu met arg glu leu glu asp gly thr pro arg ala gly
CTG GCC CTG ATG CGG GAG CTG GAA GAT GGC ACC CCC CGG GCT GGG

45
gln ile leu lys gln thr tyr asp lys phe asp thr asn met arg
CAG ATC CTC AAG CAG ACC TAT GAC AAA TTT GAC ACA AAC ATG CGC

50
ser asp asp ala leu leu lys asn tyr gly leu leu ser cys phe
AGT GAC GAC GCG CTG CTC AAG AAC TAC GGT CTG CTC TCC TGC TTC

55
arg lys asp leu his lys thr glu thr tyr leu arg val met lys
CGG AAG GAC CTG CAT AAG ACG GAG ACG TAC CTG AGG GTC ATG AAG

60
cys arg arg phe gly glu ala ser cys ala phe stop
TGC CGC CGC TTC GGG GAG GCC AGC TGC GCA TTC TAG

Detailed Description45 Synthesis of Proximal Portion of BGH Gene

Twelve fragments, U 1-6 (upper strand) and L 1-6 (lower strand), were synthesized. Also synthesized, in order to repair the 3' end of the gene, were 2 fragments, BGH Repair (1) (upper strand) and BGH Repair (2) (lower strand).

The 14 fragments were synthesized according to the method of Crea et al., *Nucleic Acids Research*, 8, 2331 (1980). The syntheses of the fragments were accomplished from the appropriate solid support (cellulose) by sequential addition of the appropriate fully protected dimer - or trimer- blocks. The cycles were carried out under the same conditions as described in the synthesis of oligothymidilic acid (see Crea et al., *Supra*.) The final polymer was treated with base (aq. conc NH₃) and acid (80% aq. HOAC), the polymer pelleted off and the supernatant evaporated to dryness. The residue, as dissolved in 4% aq. NH₃, was washed with ether (3x) and used for the isolation of the fully deprotected fragment. Purification was accomplished by hplc on Rasil NH₂ u-particulate column. Gel electrophoretic analysis showed that each of

the fragments, U,L 1-6 and BGH Repair (1) and (2), had the correct size:

	Fragment	Sequence	Size
5			
	U 1	5' AAT.TCT.ATG.TTC.C ^{3'}	13-mer
	U 2	5' CAG.CTA.TGT.CTC.T ^{3'}	13-mer
10	U 3	5' ATC.TGG.TCT.ATT.C ^{3'}	13-mer
	U 4	5' GCT.AAC.GCT.GTT.C ^{3'}	13-mer
	U 5	5' TTC.GTG.CTC.AGC.A ^{3'}	13-mer
15	U 6	5' TCT.TCA.TCA.GCT.GA ^{3'}	14-mer
	L 1	5' ATA.GCT.GGG.AAC.ATA.G ^{3'}	16-mer
	L 2	5' ACC.AGA.TAG.AGA.C ^{3'}	13-mer
20	L 3	5' CGT.TAG.CGA.ATA.G ^{3'}	13-mer
	L 4	5' GCA.CGA.AGA.ACA.G ^{3'}	13-mer
	L 5	5' ATG.AAG.ATG.CTG.A ^{3'}	13-mer
	L 6	5' AGC.TTC.AGC.TG ^{3'}	11-mer
25	BGH Repair (1)	5' AA.TTC.AGC.TGC.GCA.TTC.TAG.A ^{3'}	21-mer
	BGH Repair (2)	5' AG.CTT.CTA.GAA.TGC.GCA.GCT.G ^{3'}	21-mer

30 Construction of pBGH 33 (Fig. 4)

Fresh frozen bovine pituitaries were macerated and RNA was extracted by the guanidium thiocyanate method. (Harding et al., J. Biol Chem. 252 (20), 7391 (1977) and Ullrich et al., Science 196, 1313 (1977)). The total RNA extract was then passed over an oligo-dT cellulose column to purify poly A containing messenger RNA (mRNA). Using reverse transcriptase and oligo-dT as a primer, single stranded cDNA was made from the mRNA. Second strand synthesis was achieved by use of the Klenow fragment of DNA polymerase I. Following S1 enzyme treatment and acrylamide gel electrophoresis a size cut of the total cDNA (ca. 500-1500 bp) was eluted and cloned into the Pst I site of the amp^R gene of pBR 322 using traditional tailing and annealing conditions.

The pBR 322 plasmids containing cDNA were transformed into E. coli K-12 strain 294 (ATCC No. 31446). Colonies containing recombinant plasmids were selected by their resistance to tetracycline and sensitivity to ampicillin. Approximately 2000 of these clones were screened for BGH by colony hybridization.

45 The cDNA clones of HGH contain an internal 550 bp HaeIII fragment. The amino acid sequence of this region is very similar to the BGH amino acid sequence. This HGH HaeIII fragment was radioactively labeled and used as a probe to find the corresponding BGH sequence amongst the 2000 clones.

Eight positive clones were identified. One of these, pBGH112, was verified by sequence analysis as BGH. This full-length clone is 940 bp long containing the coding region of the 26 amino acid presequence 50 as well as the 191 amino acid protein sequence.

In order to achieve direct BGH expression, a synthetic "expression primer" was made having the sequence 5'-ATGTTCCCAGCCATG-3'. The nucleotides in the fourth through fifteenth position are identical to the codons of the first 4 amino acids of the mature BGH protein, as determined by sequence data of pBGH 112. Only the 5' ATG (methionine) is alien to this region of the protein. This was necessary in order 55 to eliminate the presequence region of our BGH clone and to provide the proper initiation codon for protein synthesis. By a series of enzymatic reactions this synthetic primer was elongated on the BGH 112 cDNA insert. The primed product was cleaved with Pst I to give a DNA fragment of 270 bp containing coding information up to amino acid 90. (Figure 4) This "expression" BGH cDNA fragment was ligated into a pBR

322 vector which contained the trp promotor. This vector was derived from pLeIF A trp25 (Goeddel et al., Nature 287, 411 (1980)). The interferon cDNA was removed and the trp25-322 vector purified by gel electrophoresis. The recombinant plasmid (pBGH710) now contained the coding information for amino acids 1-90 of the mature BGH protein, linked directly to the trp promotor. This linkage was verified by DNA sequence analysis. The second half of the coding region and the 3' untranslated region was isolated from pBGH112 by PstI restriction digest and acrylamide gel electrophoresis. This "back-end" fragment of 540 bp was then ligated into pBGH710 at the site of amino acid 90. Recombinant plasmids were checked by restriction analysis and DNA sequencing. The recombinant plasmid, pBGH33, has the trp promotor directly linked via ATG with the complete DNA coding sequence for mature BGH.

10

Construction of pHGH 207-1

Plasmid pGMI carries the *E. coli* tryptophan operon containing the deletion LE1413 (G.F. Miozzari, et al., (1978) J. Bacteriology 1457-1466)) and hence expresses a fusion protein comprising the first 6 amino acids of the trp leader and approximately the last third of the trp E polypeptide (hereinafter referred to in conjunction as LE'), as well as the trp D polypeptide in its entirety, all under the control of the trp promoter-operator system. The plasmid, 20 µg, was digested with the restriction enzyme Pvull which cleaves the plasmid at five sites. The gene fragments were next combined with EcoRI linkers (consisting of a self complementary oligonucleotide of the sequence: pCATGAATTTCATG) providing an EcoRI cleavage site for a later cloning into a plasmid containing an EcoRI site. The 20 µg of DNA fragments obtained from pGMI were treated with 10 units T₄ DNA ligase in the presence of 200 pico moles of the 5'-phosphorylated synthetic oligonucleotide pCATGAATTTCATG and in 20µl T₄ DNA ligase buffer (20mM tris, pH 7.6, 0.5 mM ATP, 10 mM MgCl₂, 5 mM dithiothreitol) at 4°C overnight. The solution was then heated 10 minutes at 70°C to inactivate ligase. The linkers were cleaved by EcoRI digestion and the fragments, now with EcoRI ends, were separated using polyacrylamide gel electrophoresis (hereinafter "PAGE") and the three largest fragments isolated from the gel by first staining with ethidium bromide, locating the fragments with ultraviolet light, and cutting from the gel the portions of interest. Each gel fragment, with 300 microliters 0.1xTBE, was placed in a dialysis bag and subjected to electrophoresis at 100 V for one hour in 0.1xTBE buffer (TBE buffer contains: 10.8 gm tris base, 5.5 gm boric acid, 0.09 gm Na₂EDTA in 1 liter H₂O). The aqueous solution was collected from the dialysis bag, phenol extracted, chloroform extracted and made 0.2 M sodium chloride, and the DNA recovered in water after ethanol precipitation. (All DNA fragment isolations hereinafter described are performed using PAGE followed by the electroelution method just discussed.) The trp promoter-operator-containing gene with EcoRI sticky ends was identified in the procedure next described, which entails the insertion of fragments into a tetracycline sensitive plasmid which, upon promoter-operator insertion, becomes tetracycline resistant.

Plasmid pBRH1, (R.I. Rodriguez, et al., Nucleic Acids Research 6, 3267-3287 [1979]) expresses ampicillin resistance and contains the gene for tetracycline resistance but, there being no associated promoter, does not express that resistance. The plasmid is accordingly tetracycline sensitive. By introducing a promoter-operator system in the EcoRI site, the plasmid can be made tetracycline resistant.

pBRH1 was digested with EcoRI and the enzyme removed by phenol/CHCl₃ extraction followed by chloroform extraction and recovered in water after ethanol precipitation. The resulting DNA molecule was, in separate reaction mixtures, combined with each of the three DNA fragments obtained as described above and ligated with T₄ DNA ligase as previously described. The DNA present in the reaction mixture was used to transform competent *E. coli* K-12 strain 294 (K. Backman et al., Proc Nat'l Acad Sci USA 73, 4174-4198 (1976) (ATCC no. 31446) by standard techniques (V. Hershfield et al., Proc Nat'l Acad Sci USA 71, 3455-3459 (1974)) and the bacteria plated on LB plates containing 20 µg/ml ampicillin and 5 µg/ml tetracycline. Several tetracycline-resistant colonies were selected, plasmid DNA isolated and the presence of the desired fragment confirmed by restriction enzyme analysis. The resulting plasmid, designed pBRHtrp, expresses β-lactamase, imparting ampicillin resistance, and it contains a DNA fragment including the trp promoter-operator and encoding a first protein comprising a fusion of the first six amino acids of the trp leader and approximately the last third of the trp E polypeptide (this polypeptide is designated LE'), and a second protein corresponding to approximately the first half of the trp D polypeptide (this polypeptide is designated D'), and a third protein coded for by the tetracycline resistance gene.

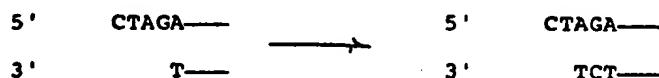
pBRH trp was digested with EcoRI restriction enzyme and the resulting fragment 1 isolated by PAGE and electroelution. EcoRI-digested plasmid pSom 11 (K. Itakura et al, Science 198, 1056 (1977); G.B. patent publication no. 2 007 676 A) was combined with this fragment 1. The mixture was ligated with T₄ DNA ligase as previously described and the resulting DNA transformed into *E. coli* K-12 strain 294 as previously described. Transformant bacteria were selected on ampicillin-containing plates. Resulting ampicillin-resistant

colonies were screened by colony hybridization (M. Gruenstein et al., Proc Nat'l Acad Sci USA 72, 3951-3965 [1975]) using as a probe the trp promoter-operator-containing fragment 1 isolated from pBRHtrp, which had been radioactively labelled with P³². Several colonies shown positive by colony hybridization were selected, plasmid DNA was isolated and the orientation of the inserted fragments determined by restriction analysis employing restriction enzymes BglII and BamHI in double digestion. E. coli 294 containing the plasmid designated pSOM7Δ2, which has the trp promoter-operator fragment in the desired orientation was grown in LB medium containing 10 µg/ml ampicillin. The cells were grown to optical density 1 (at 550 nm), collected by centrifugation and resuspended in M9 media in tenfold dilution. Cells were grown for 2-3 hours, again to optical density 1, then lysed and total cellular protein analyzed by SDS (sodium dodecyl sulfate) area (15 percent) polyacrylamide gel electrophoresis (J.V. Maizel Jr. et al., Metb Viral 5, 180-246 (1971)).

The plasmid pSom7Δ2, 10 µg, was cleaved with EcoRI and the DNA fragment 1 containing the tryptophan genetic elements was isolated by PAGE and electroelution. This fragment, 2 µg, was digested with the restriction endonuclease Taq I, 2 units, 10 minutes at 37°C such that, on the average, only one of the approximately five Tag I sites in each molecule is cleaved. This partially digested mixture of fragments was separated by PAGE and an approximately 300 base pair fragment 2 that contained one EcoRI end and one Tag I end was isolated by electroelution. The corresponding Tag I site is located between the transcription start and translation start sites and is 5 nucleotides upstream from the ATG codon of the trp leader peptide. The DNA sequence about this site is shown in Figure 4. By proceeding as described, a fragment could be isolated containing all control elements of the trp operon, i.e., promoter-operator system, transcription initiation signal, and part of the trp leader ribosome binding site.

The Tag I residue at the 3' end of the resulting fragment adjacent the translation start signal for the trp leader sequence was next converted into an XbaI site. This was done by ligating the Fragment 2 obtained above to a plasmid containing a unique (i.e., only one) EcoRI site and a unique XbaI site. For this purpose, one may employ essentially any plasmid containing, in order, a replicon, a selectable marker such as antibiotic resistance, and EcoRI, XbaI and BamHI sites. Thus, for example, an XbaI site can be introduced between the EcoRI and BamHI sites of pBR322 (F. Bolivar et al., Gene 2, 95-119 [1977]) by, e.g., cleaving at the plasmid's unique Hind III site with Hind III followed by single strand-specific nuclease digestion of the resulting sticky ends, and blunt end ligation of a self annealing double-stranded synthetic nucleotide containing the recognition site such as CCTCTAGAGG. Alternatively, naturally derived DNA fragments may be employed, as was done in the present case, that contain a single XbaI site between EcoRI and BamHI cleavage residues. Thus, an EcoRI and BamHI digestion product of the viral genome of hepatitis B was obtained by conventional means and cloned into the EcoRI and BamHI sites of plasmid pGH6 (D.V. Goeddel et al., Nature 281, 544 [1979])) to form the plasmid pHs32. Plasmid pHs32 was cleaved with XbaI, phenol extracted, chloroform extracted and ethanol precipitated. It was then treated with 1 µl E. coli polymerase I, Klenow fragment (Boehringer-Mannheim) in 30 µl polymerase buffer (50 mM potassium phosphate pH 7.4, 7mM MgCl₂, 1 mM β-mercaptoethanol) containing 0.1mM dTTP and 0.1mM dCTP for 30 minutes at 0°C then 2 hr. at 37°C. This treatment causes 2 of the 4 nucleotides complementary to the 5' protruding end of the XbaI cleavage site to be filled in:

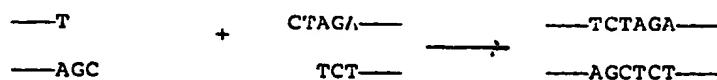
40



45

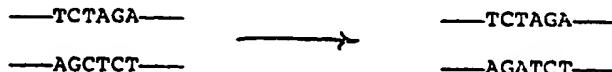
Two nucleotides, dC and dT, were incorporated giving an end with two 5' protruding nucleotides. This linear residue of plasmid pHs32 (after phenol and chloroform extraction and recovery in water after ethanol precipitation) was cleaved with EcoRI. The large plasmid fragment was separated from the smaller EcoRI-XbaI fragment by PAGE and isolated after electroelution. This DNA fragment from pHs32 (0.2 µg), was ligated, under conditions similar to those described above, to the EcoRI-Taq I fragment of the tryptophan operon (0.01 µg). In this process the Taq I protruding end is ligated to the XbaI remaining protruding end even though it is not completely Watson-Crick base-paired:

55



A portion of this ligation reaction mixture was transformed into *E. coli* 294 cells as in part I. above, heat treated and plated on LB plates containing ampicillin. Twenty-four colonies were selected, grown in 3 ml LB media, and plasmid isolated. Six of these were found to have the XbaI site regenerated via *E. coli* catalyzed DNA repair and replication:

5



10

These plasmids were also found to cleave both with EcoRI and HpaI and to give the expected restriction fragments. One plasmid 14, designated pTrp 14, was used for expression of heterologous polypeptides, as next discussed.

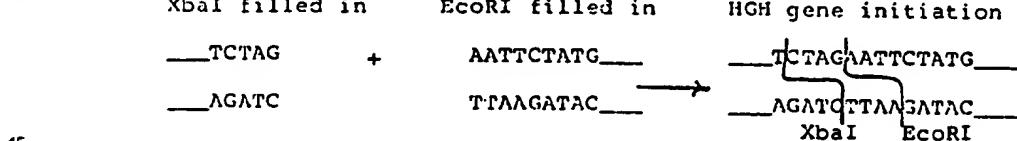
The plasmid pHGH 107 (D.V. Goeddel et al, *Nature*, 281, 544, 1979) contains a gene for human growth hormone made up of 23 amino acid codons produced from synthetic DNA fragments and 163 amino acid codons obtained from complementary DNA produced via reverse transcription of human growth hormone messenger RNA. This gene, 3, though it lacks the codons of the "pre" sequence of human growth hormone, does contain an ATG translation initiation codon. The gene was isolated from 10 µg pHGH 107 after treatment with EcoRI followed by *E. coli* polymerase I Klenow fragment and dTTP and dATP as described above. Following phenol and chloroform extraction and ethanol precipitation the plasmid was treated with BamHI.

The human growth hormone ("HGH") gene-containing fragment 3 was isolated by PAGE followed by electroelution. The resulting DNA fragment also contains the first 350 nucleotides of the tetracycline resistance structural gene, but lacks the tetracycline promoter-operator system so that, when subsequently cloned into an expression plasmid, plasmids containing the insert can be located by the restoration of tetracycline resistance. Because the EcoRI end of the fragment 3 has been filled in by the Klenow polymerase I procedure, the fragment has one blunt and one sticky end, ensuring proper orientation when later inserted into an expression plasmid.

The expression plasmid pTrp14 was next prepared to receive the HGH gene-containing fragment prepared above. Thus, pTrp14 was XbaI digested and the resulting sticky ends filled in with the Klenow polymerase I procedure employing dATP, dTTP, dGTP and dCTP. After phenol and chloroform extraction and ethanol precipitation the resulting DNA was treated with BamHI and the resulting large plasmid fragment isolated by PAGE and electroelution. The pTrp14-derived fragment had one blunt and one sticky end, permitting recombination in proper orientation with the HGH gene containing fragment 3 previously described.

The HGH gene fragment 3 and the pTrp14 Xba-BamHI fragment were combined and ligated together under conditions similar to those described above. The filled in XbaI and EcoRI ends ligated together by blunt end ligation to recreate both the XbaI and the EcoRI site:

40



45

This construction also recreates the tetracycline resistance gene. Since the plasmid pHGH 107 expresses tetracycline resistance from a promoter lying upstream from the HGH gene (the lac promoter), this construction, designated pHGH 207, permits expression of the gene for tetracycline resistance under the control of the tryptophan promoter-operator. Thus the ligation mixture was transformed into *E. coli* 294 and colonies selected on LB plates containing 5 µg/ml tetracycline.

Construction of pBGH33-1 (Figure 5)

55 The structure of pHGH207-1 which has the entire human growth hormone gene sequence is shown. The front part of this gene is synthetic as is described by Goeddel et al., *Nature* 281, 544 (1979). In the following a plasmid was constructed containing the BGH gene in the same orientation and in the same position with respect to the trp-promotor as is the HGH gene in pHGH 207-1.

Twenty μ l (i.e. 10 μ g) of the plasmid DNA was digested with Bam HI and Pvull as follows: To the twenty μ l of DNA we added 5 μ l 10X restriction enzyme buffer (500mM NaCl, 100 mM Tris HCl pH 7.4, 100 mM MgSo₄ and 10 mM DTT), 20 μ l H₂O and 10 units BamHI restriction enzyme and 2 μ l Pvull restriction enzyme.

5 Subsequently, this reaction mixture was incubated at 37°C for 90 minutes. The mixture was loaded on a 6 percent acrylamide gel and electrophoresis was carried out for 2 hours at 50 mA. The DNA in the gel was stained with Ethidium bromide and visualized with UV-light. The band corresponding to the 365 bp (with reference to a Haelll digest of pBR322) fragment was excised and inserted in a dialysis bag and the DNA was electroeluted using a current of 100 mA. The liquid was removed from the bag and its salt concentration adjusted to 0.3M NaCl. Two volumes of ethanol were added and the DNA precipitated at 10 -70°C. The DNA was spun down in an Eppendorf centrifuge, washed with 70 percent ethanol and dried and resuspended in 10 μ l TAE (10 mM Tris HCl pH7.4, 0.1 mM EDTA). Similarly, the large XbaI Bam HI fragment of pHGH 207-1 and the XbaI, partial Pvull 570 bp fragment of pBGH33 were isolated.

Two μ l of each of the thus isolated DNA fragments were mixed. 1 μ l 10mM ATP and 1 μ l 10x ligase buffer (200 mM Tris HCl pH7.5, 100mM MgCl₂, 20 mM DTT) and 1 μ l T₄ DNA ligase and 2 μ l H₂O were added. Ligation was done over night at 4°C. This mixture was used to transform competent E. coli K-12 294 cells as follows: 10 ml L-broth was inoculated with E. coli K-12 294 and incubated at 37°C in a shaker bath at 37°C. At OD₅₅₀ of 0.8 the cells were harvested by spinning in a Sorvall centrifuge for 5 min. at 6000 rpm. The cell pellet was washed/resuspended in 0.15 M NaCl, and again spun. The cells were resuspended in 75 mM CaCl₂, 5 mM MgCl₂ and 10 mM Tris HCl pH7.8 and incubated on ice for at least 20 min. The cells were spun down for 5 min at 2500 rpm and resuspended in the same buffer. To 250 μ l of this cell suspension each of the ligation mixtures was added and incubated for 60 min on ice. The cells were heat shocked for 90 seconds at 42°C, chilled and 2 ml L-broth was added. The cells were allowed to recover by incubation at 37°C for 1 hour. 100 μ l of this cell suspension was plated on appropriate plates which were subsequently incubated over night at 37°C. The plasmid structure in several of the colonies thus obtained is shown in Figure 5 (pBGH 33-1).

All further constructions were done using the same procedures, as described above, mutatis mutandis.

Construction of the hybrid growth hormone genes HBGH and BHGH (Figure 5)

30 The two Pvull sites in the HGH and BGH genes are at identical positions. Exchange of Pvull fragments is possible without changing the reading frame of the messenger RNA of these genes. The large difference in expression of both genes is due to differences in initiation of protein synthesis at the beginning of the messages. Therefore, the front part of both genes were exchanged thus constructing hybrid genes that upon transcription would give hybrid messenger RNAs. The two plasmids, pBHGH and pHBGH, were constructed as follows:

From pHGH207-1 there were isolated the large BamHI-XbaI fragment and the 857 bp BamHI (partial) Pvull fragment containing the HGH gene without its front part. From pBGH33-1 there was isolated the 75 bp XbaI-Pvull fragment that contains the front part of the BGH gene. After ligation and transformation pBHGH was obtained. pHBGH was constructed in a similar way to pBHGH; in this case the back part was derived from pBGH33-1 whereas the front part, the 75 bp XbaI-Pvull fragment, was derived from pHGH207-1.

Design and cloning of the synthetic front part of the BGH gene (Figure 6)

45 The DNA sequence up to the Pvull site of the BGH and HGH genes codes for 22 amino acids. Since the front part of the HGH gene had excellent protein synthesis initiation properties, the sequence of the front part of BGH was designed such that the number of nucleotide changes in the BGH gene would be minimal with respect to the HGH gene. Only 14 base pair changes from the natural BGH sequence were made in order to code for the proper BGH amino acid sequence and reduce conformational structure in the prospective mRNA. The DNA sequence is shown in Figure 6. The sequence ends with EcoRI and HindIII sticky ends to make cloning in a vector easy. Close to the HindIII site is a Pvull site for the proper junction with the remaining part of the BGH gene.

50 The fragments U1 to U6 and L1 to L6 were synthesized chemically according to the procedures described above. All the fragments except U1 and L6 were mixed and kinased. After addition of U1 and L6 the mixed fragments were ligated, purified on a 6 percent polyacrylamide gel and the 75 bp band extracted and isolated according to standard procedures. This fragment was inserted into pBR322 that had been cut with EcoRI and HindIII. Thus plasmid pBR322-01 was obtained.

Replacement of the natural front part of the BGH gene by the synthetic front part. (Figure 7)

From pBR322-01 the cloned synthetic front of the BGH gene was excised with EcoRI and Pvull, and the resulting 70 bp fragment was isolated. From pBGH33-1 the large EcoRI-BamHI fragment and the 875 bp 5 BamHI (partial) Pvull fragment was isolated. The three fragments were isolated and ligated and used to transform E. coli K-12 294 as described before. Thus, pBGH33-2 was obtained. This plasmid contains the entire BGH gene but does not have a promotor. Therefore, pBGH33-2 was cut with EcoRI and the trp-promotor containing 310 bp EcoRI fragment derived from pHGH207-1 was inserted by ligation. After 10 transformation tetracycline resistant colonies were analyzed. Therefore, these colonies had the inserted trp-promotor oriented towards the HGH- and tet-gene as shown in the figure.

Repair of the 3'-end of the BGH gene. (Figure 8)

The sequences beyond the second Pvull site of the BGH gene are derived from the HGH gene. One of 15 the amino acids at the end is different from that in the natural BGH gene. This 3'-end was repaired as follows. A synthetic DNA fragment as shown was synthesized. It is flanked by an EcoRI and a HindIII end to facilitate cloning and contains a Pvull site and 3 amino acid codons and a stop codon in the reading frame 20 of the BGH gene itself. This fragment was inserted into EcoRI-HindIII opened pBR322. Thus pBR322-02 was obtained. Subsequently this plasmid was cut with Pvull and BamHI and the 360 bp fragment was isolated. From pBGH33-3, which has the entire BGH gene with the synthetic front part, the large BamHI and XbaI fragment and the 570 bp XbaI (partial) Pvull fragment was isolated. These three fragments were 25 ligated and used to transform cells. Thus, pBGH33-4 was obtained. In this plasmid a unique HindIII site is present between the stop codon of the BGH gene and the start codon of the tet-mRNA. Both genes are transcribed under direction of the trp promotor.

A typical growth medium used to derepress and produce high levels of BGH per liter (Figure 9) 25 contains: 5.0 g $(\text{NH}_4)_2\text{SO}_4$, 6.0 g K_2HPO_4 , 3.0 g $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, 1.0 g sodium citrate, 2.5 g glucose, 5 mg tetracycline, 70 mg thiamine HCl, and 60 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$.

While the present invention has been described, in its preferred embodiments, with reference to the use 30 of E. coli transformants, it will be appreciated that other microorganisms can be employed mutatis mutandis. Examples of such are other E. coli organisms, e.g. E. coli B, E. coli W3110 ATCC No. 31622 (F⁻, λ-, gal⁻, prototroph), E. coli x 1776, ATCC No. 31537, E. coli D1210, E. coli RV308, ATCC No. 31608, etc., Bacillus subtilis strains, Pseudomonas strains, etc. and various yeasts, e.g., Saccharomyces cerevisiae many of which are deposited and (potentially) available from recognized depository institutions e.g., ATCC. Following the practice of this invention and the final expression of intended polypeptide product, extraction 35 and purification techniques may be those customarily employed in this art, known per se.

Claims

1. A method of improving the translational efficiency of a microbial messenger RNA encoding a heterologous functional polypeptide or a bioactive portion thereof, the method comprising:
 40 (a) determining the thermodynamic energies of regional base pairing interactions in the messenger RNA corresponding to a DNA sequence within the region extending from the transcription initiation site to nucleotide +100 of the DNA encoding the N-terminal portion of said polypeptide; and in accordance with said determination,
 (b) providing a synthetic DNA sequence characterized in that the nucleotides thereof are selected so as to provide, on transcription, corresponding messenger RNA encoding a sequence of amino acids comprising that encoded by the DNA sequence of step a) and which, by virtue of differences from the sequence of the messenger RNA referred to in step a) demonstrates reduced regional base pairing interaction leading to increased efficiency of ribosomal translation; and
 (c) ligating the DNA of step b) in proper reading frame relation with DNA encoding the C-terminal portion of said polypeptide, so as to provide DNA encoding an amino acid sequence comprising the natural sequence of said polypeptide.
 50
2. The method according to claim 1 wherein the messenger RNA of step b), within the region from nucleotide +1 to +100, is free of secondary structure having a thermodynamic energy arithmetically less than or equal to the thermodynamic energy structure formed by homologous base pairing between nucleotides 46 to 51 and nucleotides 73 to 78 of the mRNA of natural BGH as depicted in Fig. 2 hereof.
 55

3. The method according to claim 1 or claim 2 wherein the first nucleotide of said DNA sequence of step b) corresponds to nucleotide + 1 of the corresponding messenger RNA.
4. The method according to claim 1 or claim 2 wherein the first nucleotide of said DNA sequence of step 5 b) corresponds to a nucleotide of the translational start signal.
5. The method according to claim 1 or claim 2 wherein said DNA sequence of step b) extends from about the translational start signal to about 75 or more nucleotides downstream thereof.
- 10 6. The method of any one of the preceding claims wherein said heterologous functional polypeptide is bovine growth hormone.
7. The method of claim 6 wherein the bovine growth hormone lacks the BGH presequence.
- 15 8. The method according to any of the preceding claims wherein the DNA sequence of step b) is as depicted in Figure 1 as "BGH synthetic".
9. The method according to any one of the preceding claims wherein the resulting DNA sequence is inserted together with appropriately positioned translational start and stop signals into a microbial expression vector and is therein brought under the control of a microbially operable promoter, to provide the corresponding microbial expression vehicle.
- 20 10. The method according to claim 9 wherein a microorganism is transformed with said microbial expression vehicle to provide the corresponding transformed microorganism.
- 25 11. The method according to claim 10 wherein the resulting transformed microorganism is grown under suitable fermentation conditions and caused to produce said polypeptide, said polypeptide being subsequently recovered from the fermentation medium.
- 30 12. A method of producing bovine growth hormone which comprises culturing a microorganism to express DNA contained therein encoding a mature bovine growth hormone, wherein the coding sequence within the region up to nucleotide + 100 of the mRNA has been altered from that of the natural mRNA sequence of bovine growth hormone, but without altering the natural amino acid sequence, so that the resulting mRNA has conformational structure which, compared with the use of the corresponding natural bovine growth hormone coding sequence, interferes less with expression of the hormone in said microorganism.
- 35 13. A method of producing bovine growth hormone which comprises culturing a microorganism to express DNA contained therein encoding a mature bovine growth hormone, wherein the coding sequence within the about 25 N-terminal amino acids is provided by synthetic DNA whose nucleotide sequence has been altered from that of the natural nucleotide sequence of bovine growth hormone, but without altering the natural amino acid sequence, so that the resulting mRNA has conformational structure which, compared with the use of the corresponding natural bovine growth hormone coding sequence, interferes less with expression of the hormone in said microorganism.
- 40 45 14. The method of any one of claims 10 to 13 wherein the nucleotides that encode alanine within the proline-alanine-methionine sequence near the N-terminus of the hormone are GCT, rather than the GCC of the naturally occurring bovine growth hormone DNA.
- 50 15. The method of any one of claims 10 to 14 wherein said microorganism is an E.coli strain.
16. The method of any one of claims 10 to 15, wherein the bovine growth hormone is recovered and purified.

55 Patentansprüche

1. Verfahren zum Verbesseren der Translationswirksamkeit einer mikrobiellen Messenger-RNA, die für ein heterologes funktionales Polypeptid oder einen bioaktiven Abschnitt davon kodiert, wobei das Verfahren

umfaßt:

(a) das Bestimmen der thermodynamischen Energien regionaler Basenpaarungswechselwirkungen in der Messenger-RNA, die einer DNA-Sequenz innerhalb des Bereiches entspricht, der sich von der Transkriptionseinleitungsstelle zum Nukleotid + 100 der DNA erstreckt, die für den N-terminalen Abschnitt des genannten Polypeptids kodiert;

5 und gemäß der genannten Bestimmung,

(b) das Schaffen einer synthetischen DNA-Sequenz, die dadurch gekennzeichnet ist, daß die Nukleotide davon so ausgewählt sind, daß bei der Transkription entsprechende Messenger-RNA geschaffen wird, die für eine Sequenz von Aminosäuren kodiert, die jene umfassen, für welche die DNA-Sequenz von Schritt (a) kodiert, und die, dank der Unterschiede zur Sequenz der Messenger-RNA, auf die in Schritt (a) bezogenen werden, verringerte regionale Basenpaarungswechselwirkung zeigt, die zu erhöhter Wirksamkeit der ribosomalen Translation führt; und

10 (c) das Ligieren der DNA von Schritt (b) in richtige Leserasterbeziehung mit DNA, die für den C-terminalen Abschnitt des genannten Polypeptids kodiert, um DNA zu schaffen, die für eine Aminosäuresequenz kodiert, welche die natürliche Sequenz des genannten Polypeptids umfaßt.

15 2. Verfahren nach Anspruch 1, worin die Messenger-RNA von Schritt (b) innerhalb des Bereiches von Nukleotid + 1 bis + 100 frei von sekundärer Struktur ist, die eine thermodynamische Energie aufweist, die arithmetisch geringer oder gleich groß wie die thermodynamische Energiestruktur ist, die durch homologe Basenpaarung zwischen den Nukleotiden 46 bis 51 und den Nukleotiden 73 bis 78 der mRNA von natürlichem BGH, wie in Figur 2 hiervon abgebildet, gebildet wird.

20 3. Verfahren nach Anspruch 1 oder 2, worin das erste Nukleotid der genannten DNA-Sequenz von Schritt (b) dem Nukleotid + 1 der entsprechenden Messenger-RNA entspricht.

25 4. Verfahren nach Anspruch 1 oder 2, worin das erste Nukleotid der genannten DNA-Sequenz von Schritt (b) einem Nukleotid des translationalen Startsignals entspricht.

30 5. Verfahren nach Anspruch 1 oder 2, worin die genannte DNA-Sequenz von Schritt (b) sich von etwa dem translationalen Startsignal bis etwa 75 oder mehr Nukleotide stromabwärts davon erstreckt.

35 6. Verfahren nach einem der vorhergehenden Ansprüche, worin das genannte heterologe funktionale Polypeptid Rinderwachstumshormon ist.

7. Verfahren nach Anspruch 6, worin dem Rinderwachstumshormon die BGH-Präsequenz fehlt.

40 8. Verfahren nach einem der vorhergehenden Ansprüche, worin die DNA-Sequenz von Schritt (b) wie in Figur 1 als "BGH-synthetisch" dargestellt ist.

9. Verfahren nach einem der vorhergehenden Ansprüche, worin die resultierende DNA-Sequenz gemeinsam mit auf geeignete Weise angeordneten translationalen Start- und Stoppsignalen in einen mikrobiellen Expressionsvektor eingefügt wird und darin unter die Kontrolle eines mikrobiell operablen Promoters gebracht wird, um ein entsprechendes mikrobielles Expressionsvehikel zu schaffen.

45 10. Verfahren nach Anspruch 9, worin ein Mikroorganismus mit dem genannten mikrobiellen Expressionsvehikel transformiert wird, um den entsprechenden transformierten Mikroorganismus zu schaffen.

11. Verfahren nach Anspruch 10, worin der resultierende transformierte Mikroorganismus unter geeigneten Fermentationsbedingungen gezüchtet wird und dazu gebracht wird, das genannte Polypeptid zu erzeugen, wobei das genannte Polypeptid in der Folge aus dem Fermentationsmedium rückgewonnen wird.

50 12. Verfahren zur Herstellung von Rinderwachstumshormon, welches das Kultivieren eines Mikroorganismus umfaßt, um darin enthaltene DNA zu exprimieren, die für ein reifes Rinderwachstumshormon kodiert, worin die Kodierungssequenz innerhalb des Bereiches bis zu Nukleotid + 100 der mRNA von jener der natürlichen mRNA-Sequenz von Rinderwachstumshormon geändert worden ist, ohne aber die natürliche Aminosäuresequenz zu ändern, sodaß die resultierende mRNA eine Konformationsstruktur

aufweist, die, verglichen mit der Verwendung der entsprechenden natürlichen Rinderwachstumshormonkodierungssequenz, die Expression des Hormons im genannten Mikroorganismus weniger stört.

13. Verfahren zur Herstellung von Rinderwachstumshormon, welches das Kultivieren eines Mikroorganismus umfaßt, um darin enthaltene DNA zu exprimieren, die für ein reifes Rinderwachstumshormon kodiert, worin die Kodierungssequenz innerhalb der etwa 25 N-terminalen Aminosäuren durch synthetische DNA geschaffen wird, deren Nukleotidsequenz von jener der natürlichen Nukleotidsequenz von Rinderwachstumshormon geändert worden ist, ohne aber die natürliche Aminosäuresequenz zu ändern, sodaß die resultierende mRNA eine Konformationsstruktur aufweist, die, im Vergleich zur Verwendung der entsprechenden natürlichen Rinderwachstumshormonkodierungssequenz, die Expression des Hormons im genannten Mikroorganismus weniger stört.
14. Verfahren nach einem der Ansprüche 10 bis 13, worin die Nukleotide, die für Alanin innerhalb der Prolin-Alanin-Methionin-Sequenz nahe des N-Terminus des Hormons kodieren, GCT sind, und nicht die GCC der natürlich auftretenden Rinderwachstumshormon-DNA.
15. Verfahren nach einem der Ansprüche 10 bis 14, worin der genannte Mikroorganismus ein E.coli-Stamm ist.
16. Verfahren nach einem der Ansprüche 10 bis 15, worin das Rinderwachstumshormon gewonnen und gereinigt wird.

Revendications

1. Procédé pour améliorer l'efficacité de traduction d'un ARN messager microbien codant pour un polypeptide fonctionnel hétérologue ou une portion biologiquement active de ce polypeptide, procédé consistant :
 - (a) à déterminer les énergies thermodynamiques des interactions régionales d'appariement de bases dans l'ARN messager correspondant à une séquence d'ADN dans la région s'étendant du site d'initiation de transcription au nucléotide +100 de l'ADN codant pour la portion N-terminale dudit polypeptide ; et, en fonction de ladite détermination,
 - (b) à produire une séquence d'ADN synthétique caractérisée en ce que ses nucléotides sont choisis de manière à produire, par transcription, un ARN messager correspondant codant pour une séquence d'aminoacides comprenant celle codée par la séquence d'ADN de l'étape a) et qui, en raison de différences avec la séquence de l'ARN messager mentionné dans l'étape a), présente une interaction régionale réduite d'appariement de bases conduisant à une efficacité accrue de traduction ribosomale ; et
 - (c) à réunir par ligation l'ADN de l'étape b), en rapport dans un cadre de lecture convenable, avec l'ADN codant pour la portion C-terminale dudit polypeptide, de manière à produire un ADN codant pour une séquence d'aminoacides comprenant la séquence naturelle dudit polypeptide.
2. Procédé suivant la revendication 1, dans lequel l'ARN messager de l'étape b), dans la région allant du nucléotide +1 au nucléotide +100, est dépourvu de structure secondaire ayant une énergie thermodynamique arithmétiquement inférieure ou égale à l'énergie thermodynamique de la structure formée par appariement de bases homologues entre les nucléotides 46 à 51 et les nucléotides 73 à 78 de l'ARNm de la BGH naturelle, de la manière représentée sur la figure 2 de la présente invention.
3. Procédé suivant la revendication 1 ou la revendication 2, dans lequel le premier nucléotide de la séquence d'ADN de l'étape b) correspond au nucléotide +1 de l'ARN messager correspondant.
4. Procédé suivant la revendication 1 ou la revendication 2, dans lequel le premier nucléotide de la séquence d'ADN de l'étape b) correspond à un nucléotide du signal d'initiation de traduction.
5. Procédé suivant la revendication 1 ou la revendication 2, dans lequel la séquence d'ADN de l'étape b) s'étend approximativement du signal d'initiation de traduction à approximativement 75 ou plus de 75 nucléotides en aval de ce signal.

6. Procédé suivant l'une quelconque des revendications précédentes, dans lequel le polypeptide fonctionnel hétérologue est l'hormone de croissance bovine.
7. Procédé suivant la revendication 6, dans lequel l'hormone de croissance bovine est dépourvue de la
5 préséquence de BGH.
8. Procédé suivant l'une quelconque des revendications précédentes, dans lequel la séquence d'ADN de l'étape b) est conforme à celle représentée sur la figure 1 sous le nom de "BGH synthétique".
9. Procédé suivant l'une quelconque des revendications précédentes, dans lequel la séquence d'ADN
10 résultante est insérée conjointement avec un signal d'initiation et un signal de terminaison de traduction positionnés de manière appropriée dans un vecteur d'expression microbienne et est mise dans ce vecteur sous le contrôle d'un promoteur fonctionnel dans un micro-organisme, pour produire le vecteur d'expression microbienne correspondant.
15
10. Procédé suivant la revendication 9, dans lequel un micro-organisme est transformé avec le vecteur d'expression microbienne pour produire le micro-organisme transformé correspondant.
11. Procédé suivant la revendication 10, dans lequel le micro-organisme transformé résultant est cultivé
20 dans des conditions convenables de fermentation et amené à produire le polypeptide, ledit polypeptide étant ensuite séparé du milieu de fermentation.
12. Procédé de production d'hormone de croissance bovine, qui consiste à cultiver un micro-organisme pour l'expression d'un ADN présent dans ce micro-organisme codant pour une hormone de croissance
25 bovine mature, dans lequel la séquence codante dans la région allant jusqu'au nucléotide +100 de l'ARNm a été modifiée par rapport à celle de la séquence d'ARNm naturelle de l'hormone de croissance bovine, mais sans modification de la séquence d'aminoacides naturelle, de telle sorte que l'ARNm résultant possède une structure conformationnelle qui, comparativement à l'utilisation de la
30 séquence codant pour l'hormone de croissance bovine naturelle correspondante, interfère moins avec l'expression de l'hormone dans ledit micro-organisme.
13. Procédé de production d'hormone de croissance bovine, qui consiste à cultiver un micro-organisme pour l'expression de l'ADN présent dans ce micro-organisme codant pour une hormone de croissance bovine mature, dans lequel la séquence codante dans la région correspondant approximativement au
35 25 aminoacides N-terminaux est produite par un ADN synthétique dont la séquence de nucléotides a été modifiée par rapport à la séquence de nucléotides naturelle de l'hormone de croissance bovine, mais sans modifier la séquence d'aminoacides naturelle, de telle sorte que l'ARNm résultant possède une structure conformationnelle qui, comparativement à l'utilisation de la séquence codant pour l'hormone de croissance bovine naturelle correspondante, interfère moins avec l'expression de l'hormone
40 dans ledit micro-organisme.
14. Procédé suivant l'une quelconque des revendications 10 à 13, dans lequel les nucléotides qui codent pour lalanine dans la séquence proline-alanine-méthionine à proximité de l'extrémité N-terminale de l'hormone sont les nucléotides GCT, au lieu des nucléotides GCC de l'ADN d'hormone de croissance bovine naturelle.
45
15. Procédé suivant l'une quelconque des revendications 10 à 14, dans lequel le micro-organisme est une souche de E. coli.
- 50 16. Procédé suivant l'une quelconque des revendications 10 à 15, dans lequel l'hormone de croissance bovine est séparée et purifiée.

BGH natural { amino acids: Met Phe Pro Ala Met Ser Leu Ser Gly Leu Phe Ala Asn Ala Gln His Leu His Gln
 bases: ATG TTC CCA GCC ATG TCC TTC GCG CTG AAC GCT GTC CTC CAG GCT CAC CTG CAT CAG PvuII

HGH synthetic {
 amino acids: Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln
 bases: ATG TTC CCA ACT ATA CCA CTA TCT CGT CTA TTC GAT AAC GCT ATG CTT CGT CAT GCT CTT CAT CAT CAG PvUII

BGH synthetic {
 amino acids: Met Phe Pro Ala Met Ser Leu Ser Gly Leu Phe Ala Val Asn Ala Gln His Leu His Gln
 bases: ATG TTC CCA GCT ATG TCT CTA TCT GCT AAC GCT GTT CTI CGT GCT CAG CAI CTI CAT CAG Pvull

FIG. I.

Synthetic BGH:

Natural BGH:

Restriction enzyme cleavage sites and sequencing gel analysis results are shown for Synthetic and Natural BGH. The synthetic sequence is: +1 AAGTTCTCGTAAAAAGGGTATCTAGATGTTCCCAGCCATGTCCCTG-CTCAGGCCCTGCATCAG. The natural sequence is: +1 AAGTTCTCGTAAAAAGGGTATCTAGAATTCTAGAATTCATGTTCCCAGCTATGTCTCTA CTCAGCATCTTCATCAG.

Sequencing gel analysis shows the following results:

- PvuII:** Synthetic BGH: Lanes 60, 70, 80, 90. Natural BGH: Lanes 50, 60, 70, 80, 90.
- XbaI:** Synthetic BGH: Lanes 10, 20. Natural BGH: Lanes 30, 40.
- G-C:** Synthetic BGH: Lane 50. Natural BGH: Lanes 50, 60, 70, 80, 90.

Natural BGH				Synthetic BGH			
<u>5'</u>	<u>3'</u>	<u>length</u>	<u>kcal/mol</u>	<u>5'</u>	<u>3'</u>	<u>length</u>	<u>kcal/mol</u>
14	45	8	-11.80	14	79	7	-5.50
16	31	6	4.00	16	37	6	-4.00
33	101	6	-15.40	38	104	6	-15.40
46	78	6	-15.20	52	84	-	>-10

FIG. 3.

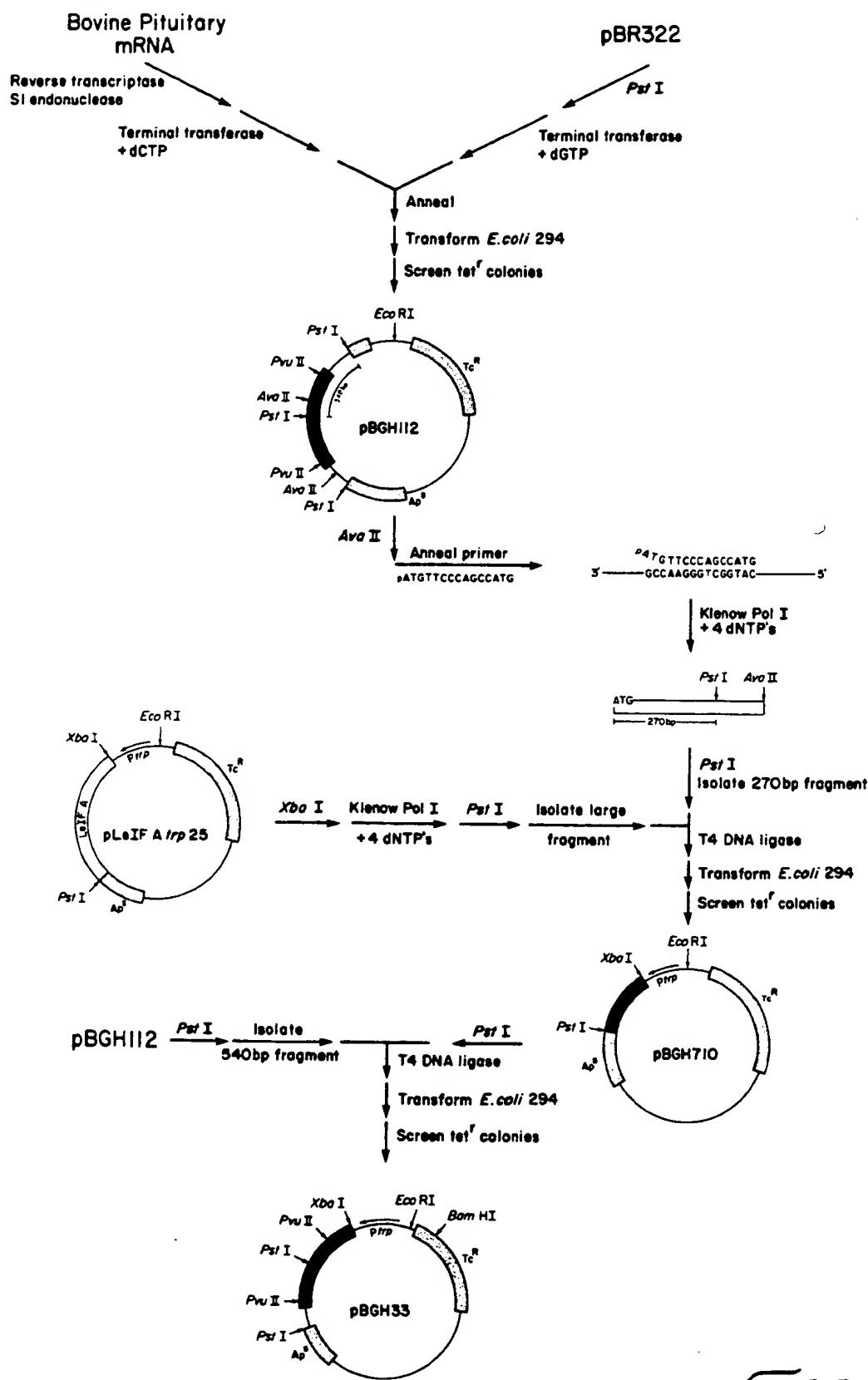


FIG. 4.

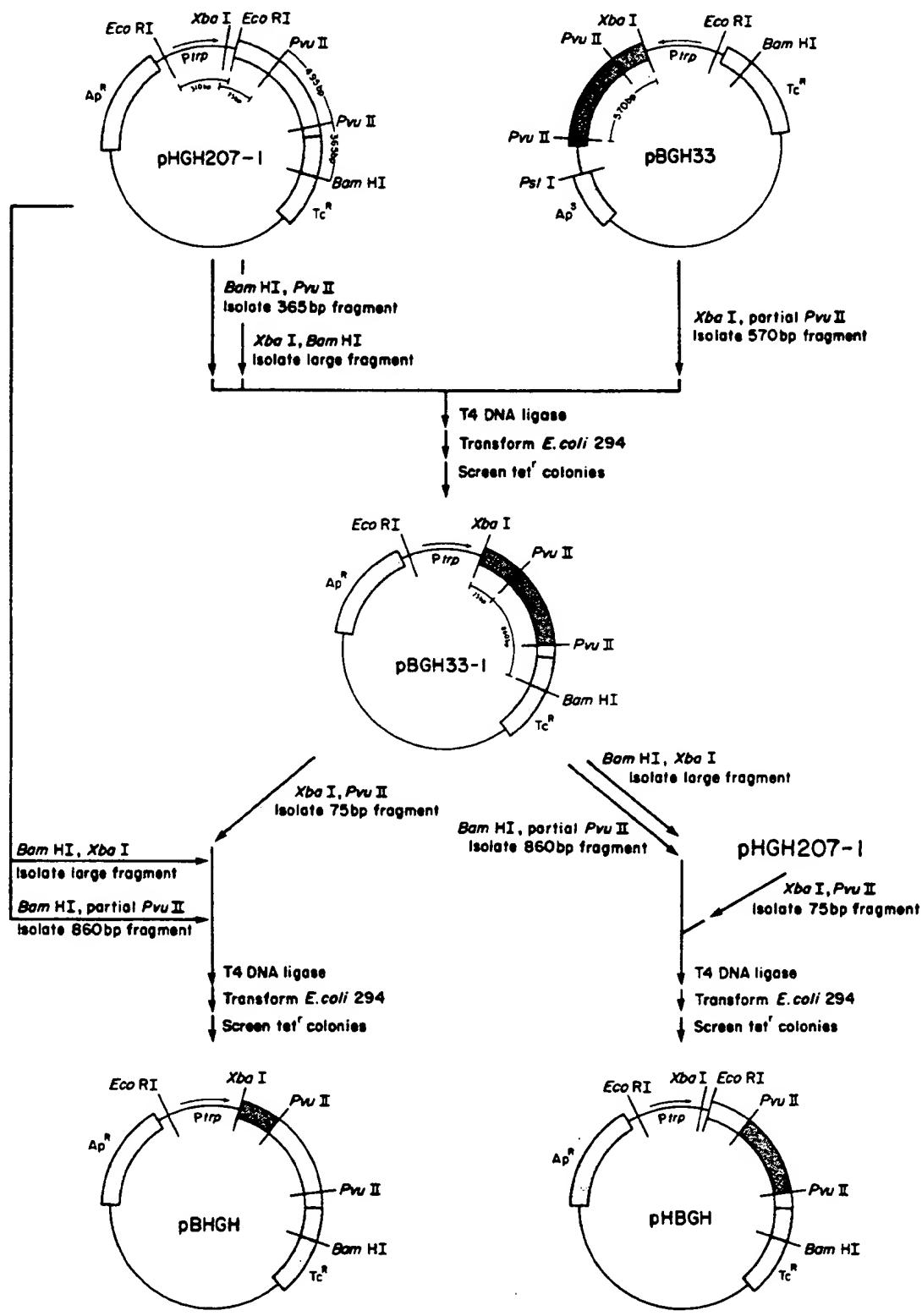
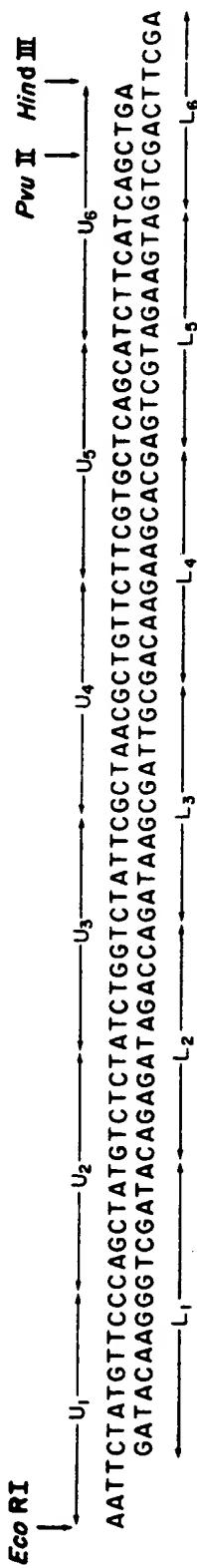


FIG. 5.



Mix Synthetic Oligonucleotides

$U_2 - U_3 - U_4 - U_5 - U_6$
 $L_1 - L_2 - L_3 - L_4 - L_5$

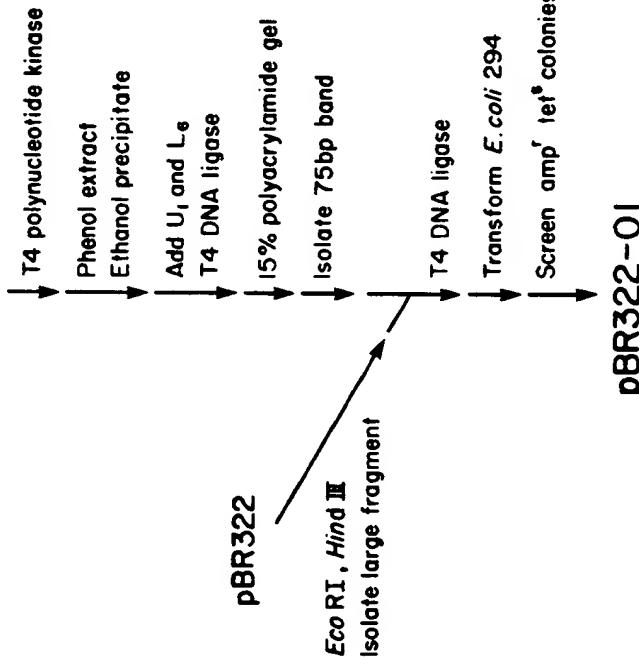


FIG. 6.

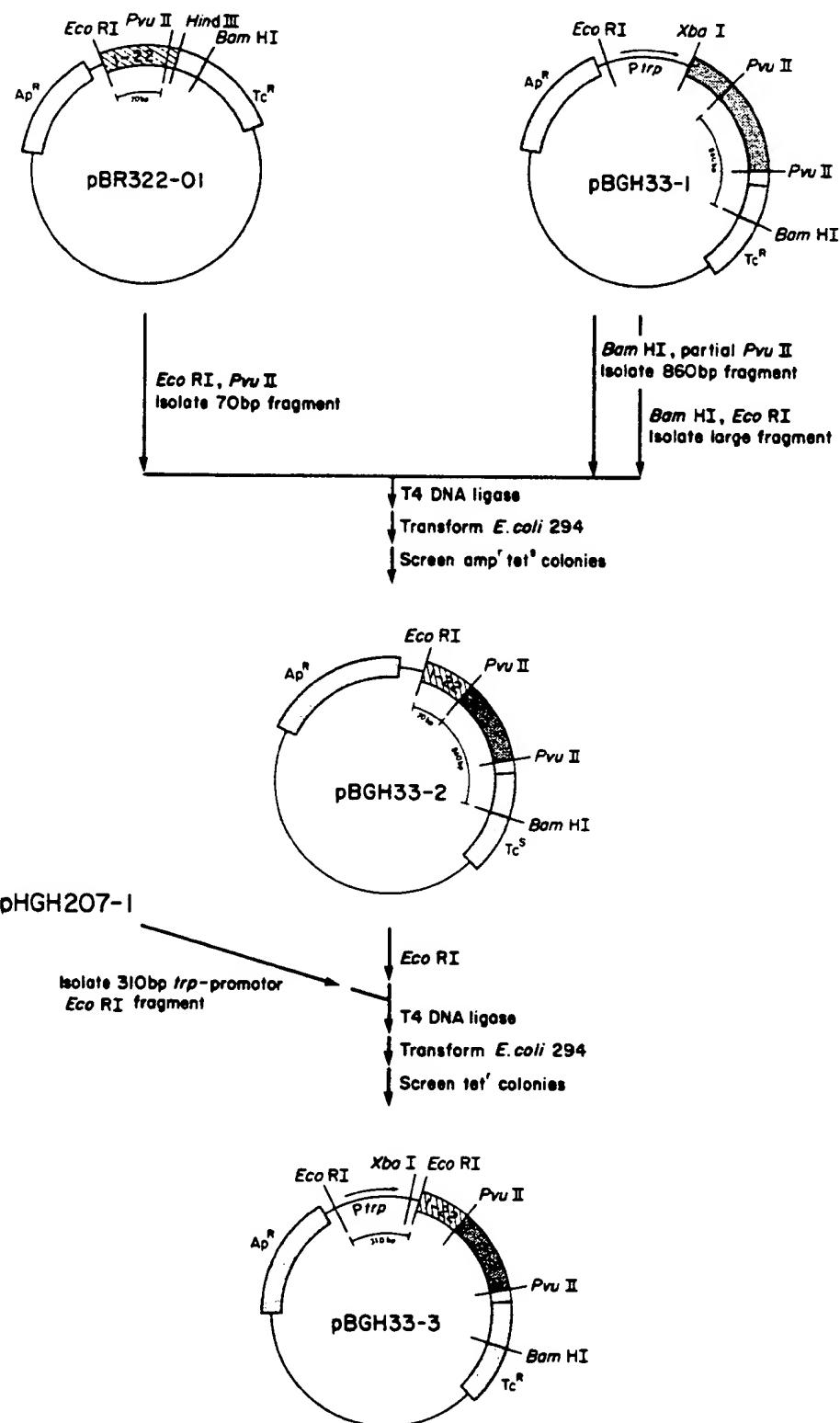


FIG. 7

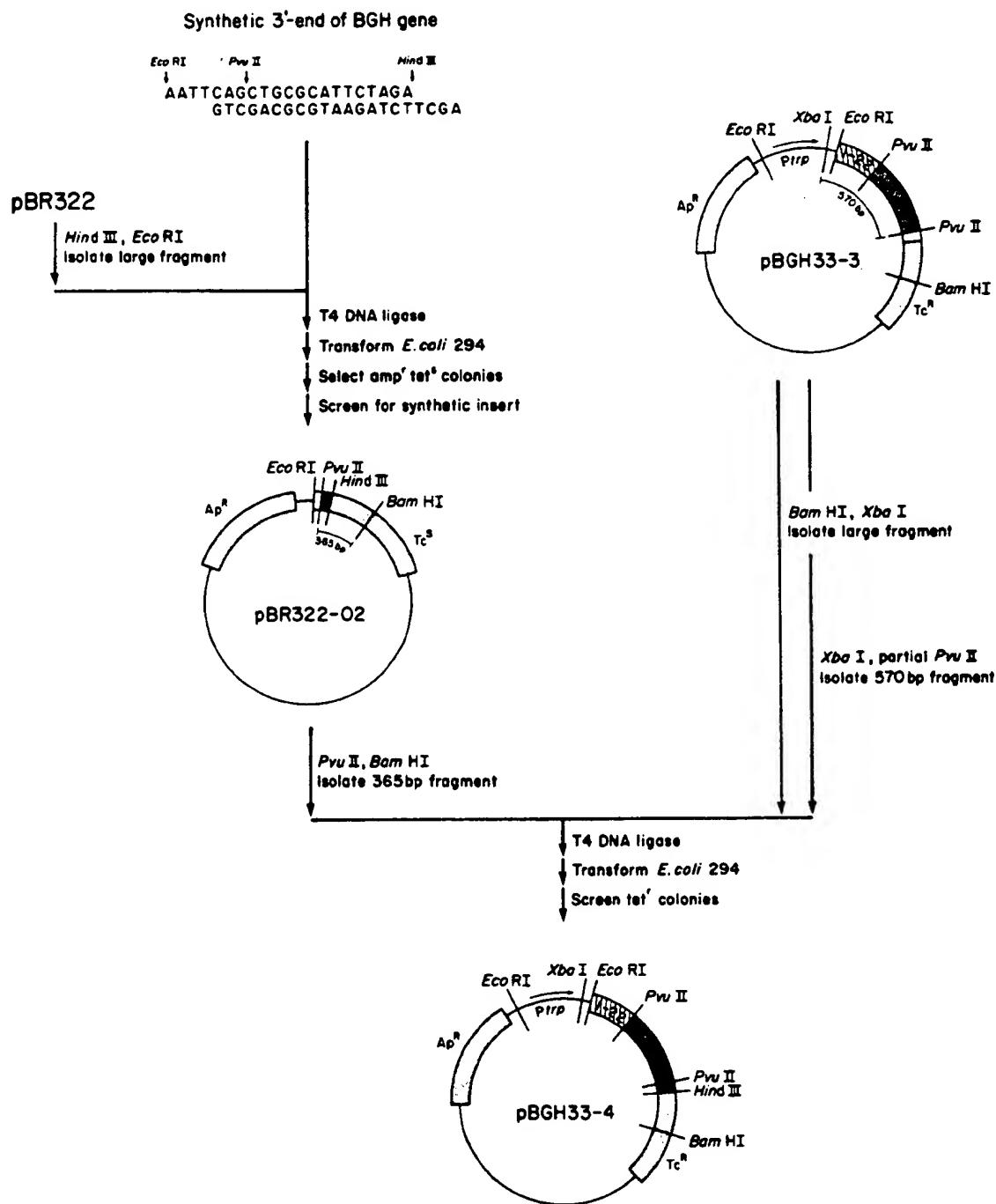


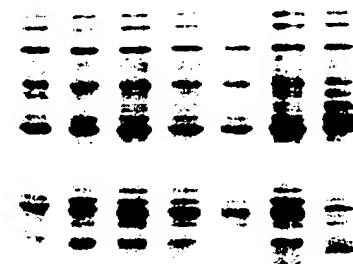
FIG. 8.

A



2.1 | 4.1 | 7.3 | 9.8 | 15.8 | GENENTECH
bGH | bGH | bGH | bGH | hGH
bGH

B



hGH | bGH | bGH | (+) | (-) |
bGH #1 bGH #2 hGH (-) hGH
#2

FIG. 9.